**CIS 4640**

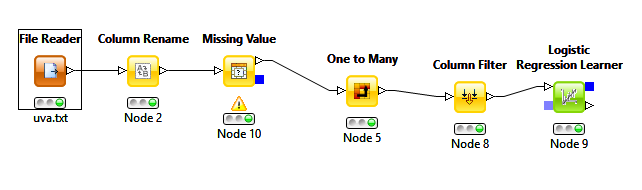
**Assignment 3**

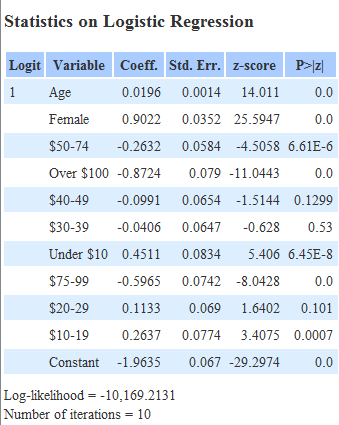
**2015-11-12**

**Todd Urbanowicz**

# Section 1 – Logistic Regression

4. a. Multi-collinearity





## Regression Equation before Testing

O = -1.9635 + 0.0196\*Age + 0.9022\*Female – 0.2632\*$50-74 – 0.8725\*Over $100 – 0.0991\*$40-49 – 0.0406\*$30-39 + 0.4511\*Under $10 – 0.5965\*$75-99 + 0.1133\*$20-29 + 0.2637\*$10-19

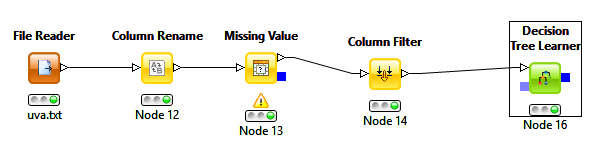
## After Hypothesis Testing

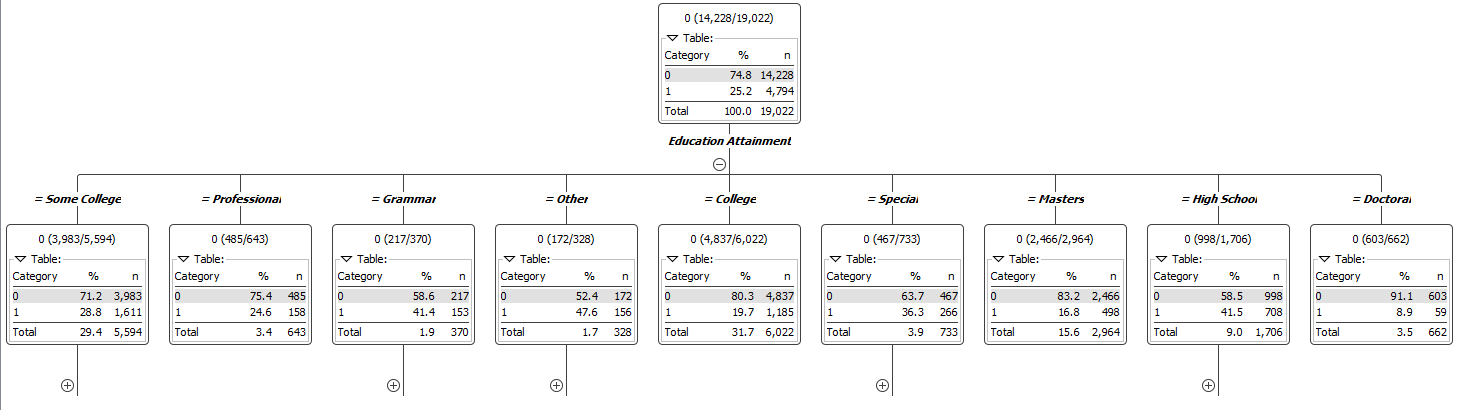
O = -1.9635 + 0.0196\*Age + 0.9022\*Female – 0.2632\*$50-74 – 0.8725\*Over $100 + 0.4511\*Under $10 – 0.5965\*$75-99 + 0.2637\*$10-19

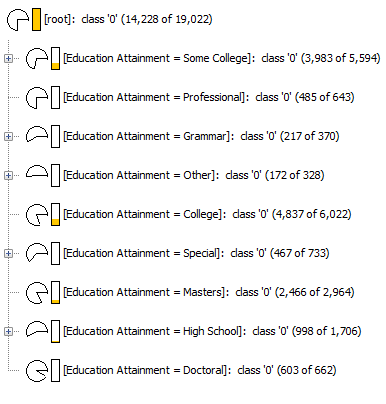
## Questions

* Q1
  + Top three predictors
    - Female, Over $100, $75-99
  + They have the largest coefficients and therefore affect the result the most
* Q2
  + Numerator 0.15966159
  + Denominator 1.15966159
  + Prob(noobie=yes|age=20|male=yes|$50000 income) 0.137679467
* Q3
  + Target females of all ages that have a household income above $75,000

# Section 2 – Classification Tree







## Questions

* Q1
  + Top Predictors
    - Education Attainment, Gender, Income
* Q2
  + No, there is only a 30.8% chance that she is a newbie.
* Q3
  + We see that the highest number of internet newbies occur in people that have no college level education, but have had grammar, high school, or other type of education. That said, in those categories, females tend to take the upper hand in being an internet newbie in all three. Almost 60% of females in all categories are internet newbies. For high school educated females, the highest percentage of internet newbies occur with a marital status of married, single, divorced, or N/A. We cannot use N/A for a target market so my suggestion is to target females without college experience that have a marital status of married, single, or divorced.